Marked point process hotspot maps for homicide and gun crime prediction in Chicago

George Mohler

Department of Mathematics and Computer Science
Santa Clara University

Abstract

Crime hotspot maps are a widely used and successful method for displaying spatial crime patterns and allocating police resources. However, hotspot maps are often created over a single timescale using one crime type. In the case of short-term hotspot maps that utilize several weeks of crime data, risk estimates suffer from high variance, especially for low frequency crimes such as homicide. Long-term hotspot maps that utilize several years of data fail to take into account near-repeat effects and emerging hotspot trends. In this paper we show how point process models of crime can be extended to include leading indicator crime types, while capturing both short-term and long-term patterns of risk, through a marked point process approach. Several years of data and many different crime types are systematically combined to yield accurate hotspot maps that can be used for the purpose of predictive policing gun-related crime. We apply the methodology to a large, open source data set made available to the general public online by the Chicago Police Department.

1. Introduction

In this article we develop a methodology for the prediction of homicide, along with precursory gun crimes, with application to predictive policing. The problem we consider is as follows: given all past homicide events (time and location), and all other gun-related crime reports in a police records management system (RMS), rank a list of geographic areas in a city according to the risk of homicide. We use homicide and gun crime in Chicago for illustration, but our methodology applies more generally to other crime types and their leading indicators.

Similar problems are considered in a number of studies (Bowers et al., 2004; Chainey et al., 2008; Mohler et al., 2011; Kennedy et al., 2010; Wang and Brown, 2012; Wang et al., 2012; Liu and Brown, 2003; Weisburd et al., 2012) and their solution can be used to allocate patrol resources each day according to risk. Algorithms typically fall into one of two broad categories, namely nonparametric methods utilizing only event data (kernel hotspot maps being the predominant choice) or multivariate models that explicitly incorporate additional variables such as demographics (Wang et al., 2012), income levels (Liu and Brown, 2003), distance from crime attractors (Wang et al., 2012; Liu and Brown, 2003; Kennedy et al., 2010), and leading-indicator crimes (Cohen et al., 2007; Gorr, 2009). In multivariate models of hotspots, static variables such as demographics and distance to crime attractors are predictive of long term crime hotspots, whereas recent event activity is predictive of short term hotspots. In the case of kernel density estimation (KDE) and other event based approaches, multiple timescales are usually not reflected in models and either long term hotspot maps are created using several years of data (Weisburd et al., 2012) or short term hotspots maps are created using several weeks or months of data and a spatial bandwidth on the order of tens or hundreds of meters (Chainey et al., 2008; Kennedy et al., 2010; Wang and Brown, 2012). One of the goals of this article is to illustrate how event-based models of hotspots can be constructed to estimate both short and long term hotspots in a systematic way.

In recent articles focused on multivariate modeling (Wang and Brown, 2012; Kennedy et al., 2010), short term hotspot maps serve as a method for accuracy comparison due to their widespread operational use. Here the short term hotspot map serves as a straw man, as low event counts lead to high variance in risk assessment. Multivariate models, on the other hand, are well suited to handle low event counts by reducing variance with the introduction of spatial variables correlated with crime rates. However, police departments' RMS often contain crime reports going back years, if not decades, thus containing large data sets that can be used to reduce variance and significantly improve the accuracy of crime hotspot maps. Here we believe the benefits of hotspot maps stand out, as for high event counts 1) the error due to variance is significantly reduced, 2) nonparametric estimates are potentially less biased than multivariate models, and 3) kernel based hotspot maps facilitate predictive analytics software that is robust and portable across various agencies without the need to gather data outside of the RMS.

We extend the point process model of burglary introduced in (Mohler et al., 2011) to a marked point process that allows for several years of crime data, and multiple crime types, to be utilized by hotspot maps. The model incorporates both fixed risk heterogeneity across the city and temporally dynamic risk. Whereas the model in (Mohler et al., 2011) is fully non-parametric, we consider a parametric version of the triggering kernel to balance the added computational cost associated with the incorporation of leading indicator crimes.

Our model is related to the decomposition of hotspots in (Gorr and Lee, 2012) into chronic and temporary hotspots. Chronic hotspots are long term in duration and necessitate problem oriented policing strategies to address the root causes of crime (Clarke and Eck, 2005; Weisburd et al., 2012). Chronic hotspots, defined by high crime volume over several years rather than several weeks, can capture a large percentage of crime within a small percentage of the area of a city (Weisburd et al., 2012). Temporary hotspots, on the other hand, last on the time scale of weeks or months. Models and policing strategies thus must be able to detect and react to emerging trends in order to deter temporary hotspots, otherwise the hotspots may have moved before maps are generated and patrols deployed. For example, the Los Angeles Police Department updates predictive policing models for every 8 hour shift as new crimes are reported and directs patrols accordingly.

One technical issue that arises in the creation of kernel hotspot maps is the selection of bandwidth or "search radius." In this work an Expectation-Maximization (EM) algorithm is developed that allows for the automated selection of model parameters, avoiding the need for hotspot bandwidth parameters to be manually tuned by the crime analyst. We apply the methodology to gun crime and homicide data in Chicago, illustrating that the predictive accuracy of hotspot maps is significantly improved when utilizing large data sets over several years. In particular, we find that dynamic hotspots account for only 12 to 15% of gun crime in Chicago and that chronic hotspots are the most dominant component of the estimated point process model. The outline of the article is as follows. In Section 2, we briefly review the mathematics of hotspot maps and we draw the connections between kernel hotspot maps, self-exciting point processes, and mixture models. In Section 3, we provide details of the marked point process model and in Section 4 we provide details on the EM algorithm for estimating model parameters. In Section 5, we provide results on the application of the model to homicide prediction in Chicago.

2. Hotspot maps, self-exciting point processes, and mixture models

2.1. Hotspot maps

Crime hotspot maps are a widely used method for visualizing spatial crime patterns, where spatial maps are color coded based upon levels of criminal activity. Our focus here is on kernel based hotspot maps, which exhibit the important features of hotspot mapping, and we refer the reader to (Chainey and Ratcliffe, 2005) for a more comprehensive treatment of the subject.

Given a spatial-temporal crime data set of event locations (x_i, y_i) and times t_i, a common method of constructing a kernel hotspot map is to use a subset of the data consisting of all crimes occurring within a specified time interval $[T_1, T_2]$:

$$\lambda(x, y) = \sum_{t_i \in [T_1, T_2]} g(x - x_i, y - y_i) \quad (1)$$

In practice this time window is often chosen to be the past several weeks or months leading up to the present, but could also be years (Gorr and Lee, 2012) to estimate chronic hotspots. The kernel g is often a 2D function that decays from the origin.

While the kernels are defined irrespective of a discretization of space, evaluating $\lambda(x, y)$ on a grid is still necessary for visualization. Furthermore, hotspot maps can be used to flag high crime areas for policing intervention (Bowers et al., 2004; Chainey et al., 2008), in which case the values of $\lambda(x, y)$ in different discrete regions of a city are used to rank those areas in terms of priority for receiving policing attention.

From a mathematical perspective, a hotspot map can be viewed as a nonparametric estimate of a stationary Poisson process over the time interval $[T_1, T_2]$. Furthermore, by taking the time interval to be moving, $[t-T, t]$, with t the present day, hotspot maps can also function as a nonparametric estimator of a nonstationary point process. For this purpose a more general "prospective" hotspot map has been introduced (Bowers et al., 2004),

$$\lambda(x, y, t) = \sum_{t > t_i} g(x - x_i, y - y_i, t - t_i), \quad (2)$$

where g is a space-time kernel that allows more recent events to be weighted higher than events further in the past. Research on near-repeat offender behavior (Bowers et al., 2004; Farrell and Pease, 2001; Johnson, 2008; Johnson et al., 2007) suggests that such an approach is useful in a policing context for responding to the contagion effects associated with certain crime types.

There is an inherent drawback to (2), however, in that the temporal decay of g creates a small sample size problem. For example, if the timescale over which g decays is on the order of several weeks, then only several weeks of data will be used to assess crime risk each day. In a typical city this may result in only several hundred events being used, when several thousand (or tens of thousand) may be needed to reduce the variance of the estimator low enough to achieve high accuracy in risk assessment. The same issue arises for (1) if the time interval is chosen to be several weeks or months, a common interval length used in police departments. The alternative is then to use more data, i.e. choosing $[T_1, T_2]$ to encompass the entire crime data set. The disadvantage here is that all crimes in (1) are weighted equally and thus the map cannot respond to the near-repeat trends present in crime data.

2.2. *Self-exciting point processes*

To address these issues, self-exciting point processes have been introduced (Mohler et al., 2011), taking the form,

$$\lambda(x, y, t) = \mu(x, y) + \sum_{t > t_i} g(x - x_i, y - y_i, t - t_i). \quad (3)$$

The advantage of this formulation of a hotspot map is that the rate of crime is decomposed into a stationary component, μ, and a component analogous to the prospective hotspot map that captures near-repeat effects. When discretized on a grid, μ models chronic differences in crime rates across different areas of the city, similar to the "chronic" hotspots described in (Gorr and Lee, 2012). Because crime records in a typical RMS often go back 10 years or more, large data sets can be used to achieve a high level of accuracy when estimating μ. Furthermore, state of the art optimization methods (Murphy, 2012; Veen and Schoenberg, 2008) can be employed to estimate the parameters of g determining the spatial and temporal length scales of near-repeat risk.

2.3. Mixture models

To conclude the section, we make the connection between hotspot maps, self-exciting point processes, and kernel mixture models. First, the background rate μ in (3) needs to be estimated from crime data (x_i, y_i, t_i). Because μ represents a stationary Poisson process, one way to estimate μ is to use KDE:

$$\mu(x, y) = \alpha \sum_{t > t_i} g_2(x - x_i, y - y_i), \qquad (4)$$

where α is a parameter to be estimated and g_2 is a kernel, possibly different from g. Combining (3) and (4) yields,

$$\lambda(x, y, t) = \alpha \sum_{t > t_i} g_2(x - x_i, y - y_i) + \sum_{t > t_i} g(x - x_i, y - y_i, t - t_i). \qquad (5)$$

Equation (5) can be viewed as a 3D mixture model where the mixtures are centered at the crime locations and the number of mixtures is equal to twice the number of events in the data set.

We note that (4) is not consistent in the sense that only one component of the process is assumed stationary. Thus, if the events follow a self-exciting point process, some events are triggered by previous events and should not be used to estimate μ. In (Mohler et al., 2011; Zhuang et al., 2002), a probabilistic approach, stochastic declustering, is used in place of (4). However, we prefer the use of (5) for two reasons: 1) it makes explicit the connection between self-exciting point processes and kernel based hotspot maps and 2) it facilitates a robust, efficient inference procedure to be detailed in the next section.

3. Marked point process model of homicide

In many instances there may be little difference in the situation and intent separating an assault with a deadly weapon and a homicide. The occurrence of serious violent crimes may provide as much, or more, information on the rate of homicide as actual homicides. We therefore take the following marked point process approach to modeling the intensity of homicides. Given categorical marks M, $M = 1, ..., N_c$, representing N_c crime types believed to be precursory to homicide (with homicide being marked by $M = 1$), the intensity of homicide is modeled as,

$$\lambda(x, y, t) = \mu(x, y) + \sum_{t > t_i} g(x - x_i, y - y_i, t - t_i, M_i). \qquad (6)$$

For the triggering kernel, g is specified as exponential in time, Gaussian in space:

$$g(x, y, t, M) = \theta(M)\omega \exp\big(-\omega t\big) \times \frac{1}{2\pi\sigma^2} \exp\bigg(-(x^2 + y^2)/(2\sigma^2)\bigg). \qquad (7)$$

Here ω determines the timescale over which contagion effects decay, σ determines the length scale over which near-repeat risk extends in space, and $\theta(M)$ determines the average number of events triggered by an initial event of type M. The choice of exponential and Gaussian functions allows for weighted sample mean estimators to be used in the maximization step of the EM algorithm for parameter estimation. We allow only the productivity parameter $\theta(M)$ to depend on crime type to reduce the overall number of model parameters. We then use all crime types considered in this paper to estimate the background rate,

$$\mu(x, y) = \sum_{t > t_i} \frac{\alpha(M_i)}{T} \frac{1}{2\pi\eta^2} \times \exp\bigg(-\big((x - x_i)^2 + (y - y_i)^2\big)/(2\eta^2)\bigg), \qquad (8)$$

where T is the length of the time window of the dataset up to the present day and $\alpha(M)$ determines the contribution of an event of type M to the background rate. Overall there are $2 \cdot N_c + 3$ parameters to be estimated: $\bigg(\omega, \sigma, \eta, \theta(M), \alpha(M)\bigg)$.

4. Expectation-Maximization algorithm for parameter estimation

The mixture model specified by Equations (7)-(8) is fit using an Expectation-Maximization algorithm

(Murphy, 2012; Veen and Schoenberg, 2008; Mohler et al., 2011). Roughly speaking, each crime in the data set is assumed to be generated by one of the mixture kernels. However, this information is unobservable, so a probability is assigned to the crime being generated by each of the mixtures. The algorithm converges when these probabilities are proportional to the value of the kernel at the crime space-time location relative to the sum of all of the kernels at the crime location (i.e. the intensity λ). Below we state the necessary equations for applying the EM algorithm, see (Murphy, 2012) for a derivation of the EM algorithm for Gaussian mixtures or (Veen and Schoenberg, 2008) for a derivation of the EM algorithm for parametric self-exciting point processes.

Let $K = K_1 + ... + K_{N_c}$ be the total number of crimes (of all types) in the dataset and K_1 be the number of homicides. Along the lines of (Mohler et al., 2011), we introduce matrices p_{ij} and p^b_{ij} of size $K \times K_1$ containing the probabilities that event i triggered homicide j through either the triggering kernel g or the background rate kernel respectively. Thus we have the E-step equations:

$$p_{ij} = \theta(M_i)\omega \exp\big(-\omega(t_j - t_i)\big)\frac{1}{2\pi\sigma^2} \times \exp\left(-\frac{(x_j - x_i)^2 + (y_j - y_i)^2}{2\sigma^2}\right)/\lambda(x_j, y_j, t_j), \quad (9)$$

$$p^b_{ij} = \frac{\alpha(M_i)\frac{1}{2\pi\eta^2}\exp\left(-\frac{(x_j-x_i)^2+(y_j-y_i)^2}{2\eta^2}\right)}{T\lambda(x_j, y_j, t_j)}. \quad (10)$$

Note that $p_{ij} > 0$ only if $t_j > t_i$, as the triggering kernel is one sided in time. Furthermore, to prevent point masses from forming during the EM algorithm, the origin must be deleted from both the triggering and background rate kernels. Thus $p_{ij} = 0$ and $p^b_{ij} = 0$ if events i and j occur at the same location (j indexes the homicides, which are a subset of events $i = 1, ..., K$).

Given p_{ij} and p^b_{ij}, the M-step consists of updating the parameters as follows:

$$\omega = \frac{\sum_{i=1}^{K}\sum_{j=1}^{K_1} p_{ij}}{\sum_{i=1}^{K}\sum_{j=1}^{K_1} p_{ij}(t_j - t_i) + \sum_{i=1}^{K}\theta(M_i)(T - t_i)e^{-\omega(T-t_i)}}, \quad (11)$$

$$\sigma^2 = \frac{\sum_{i=1}^{K}\sum_{j=1}^{K_1} p_{ij}\big((x_j - x_i)^2 + (y_j - y_i)^2\big)}{2\sum_{i=1}^{K}\sum_{j=1}^{K_1} p_{ij}}, \quad (12)$$

$$\eta^2 = \frac{\sum_{i=1}^{K}\sum_{j=1}^{K_1} p^b_{ij}\big((x_j - x_i)^2 + (y_j - y_i)^2\big)}{2\sum_{i=1}^{K}\sum_{j=1}^{K_1} p^b_{ij}}, \quad (13)$$

$$\theta(L) = \frac{\sum_{i=1}^{K}\sum_{j=1}^{K_1} p_{ij}1\{M_i = L\}}{K_L - \sum_{i=1}^{K} e^{-\omega(T-t_i)}1\{M_i = L\}}, \quad (14)$$

and

$$\alpha(L) = \frac{\sum_{i=1}^{K}\sum_{j=1}^{K_1} p^b_{ij}1\{M_i = L\}}{K_L}. \quad (15)$$

Given an initial guess for the parameters $\Big(\omega, \sigma, \eta, \theta(M), \alpha(M)\Big)$, the EM algorithm alternates between updating the probabilities in Equations (9)-(10) and then updating the parameters using equations (11)-(15). In Equation (11) and (14) we included temporal boundary correction terms (second term in the denominator). When the time window over which estimation occurs is much larger than the timescale over which near-repeat effects occur, these terms can be ignored and the equations simplify to weighted sample mean estimators. When these corrections are included, we find empirically that the previous guess for ω can be used on the right hand side. Given the large number of parameters, the EM algorithm has the potential to converge to different local minima depending on the initial parameter value guess. We find empirically that this possibility is greatly reduced by forcing the spatial bandwidths to be equal, i.e. $\sigma = \eta$. This is accomplished by removing Equation (13) and summing the numerators and denominators in (12)-(13). For the data sets used in this paper, we find that the EM algorithm converges after two hundred iterations. The computational cost of the EM algorithm is on the order of the product of the number of iterations, total number of crimes, and the number of homicides. For the dataset used in this study, the computationally complexity is $O(10^2 \times 10^4 \times 10^3)$ and training model parameters over three years of data for all of Chicago takes on the order of one hour on a laptop computer. However, model parameters need only be computed once every few weeks or months and a hotspot map can be created in a few seconds. Furthermore, in cloud predictive policing architectures, model training can be done in the background in the cloud and hotspot maps can be precomputed and stored, making the creation of EM generated hotspot maps appear instantaneous to the user. Another way to reduce the computational cost is to perform inter-point computations only on crimes within a certain geographic distance of each other, for example a cutoff on the order of 10km (much larger than the spatial bandwidth).

5. Application to Chicago gun crime and homicide

We apply the marked point process methodology to an open source data set consisting of 78,852 violent crimes occurring in Chicago, Illinois in the years 2007 through 2012. In total there are 2,803 homicides ($M = 1$) and the following gun related crimes with "handgun" in the description field: 34,637 robberies ($M = 2$), 13,888 assault ($M = 3$), 16,187 weapons violations ($M = 4$), 10,725 batteries ($M = 5$), and 612 criminal sexual assault ($M = 6$). The data can be downloaded from the website "https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2".

The application we consider is that of short term resource allocation. Hotspot policing consists of first ranking geographic locations in a city by crime counts or estimates of risk over a historical observation window, typically several months or years, and then allocating patrols or other policing interventions to the highest ranked geographic areas over an experimental period, typically several months (Braga, 2001). The number of hotspots is chosen based upon the amount of police resources available. Predictive policing in contrast consists of ranking hotspots on a daily basis according to estimated risk using both recent and historical crime incident data, rather than selecting a fixed set of hotspots for a several month intervention period.

An empirical accuracy measure of a hotspot ranking method is to measure the amount of crime that would have fallen within the hotspots flagged for intervention over the experimental period in the absence of police. In practice this is accomplished through a retrospective analysis (Bowers et al., 2004; Mohler et al., 2011), by simulating experimental periods in historical data and comparing the volume of crime captured by each ranking method. Methods can then be compared not only for a fixed number of hotspots, but for varying numbers of flagged hotspots as resources may increase or decrease. This metric is similar to a ROC curve (Gorr, 2009), but differs because of the introduction of the daily resource constraint in place of using a varying alarm threshold and associated false positive rate on the x-axis.

In particular, we divide Chicago into 150m$\times$150m cells that will serve as the potential hotspots to be ranked each day. We then simulate three year-long hotspot policing experiments in the years 2010, 2011, and 2012. For year 2010, we train the marked point process model parameters over the three year period 2007-2009 and then use these parameters throughout 2010. Each day in 2010, the intensity of the point process is used to rank the hotspot grid cells across Chicago. It should be noted that the dynamic component of the point process requires as input recent crimes in the history of the process. Thus to rank hotspots on day d in 2010 we use all crimes in the history of the process through day $d - 1$. We then calculate the percentage of crime $f(x)$ falling in the top $x\%$ of the ranked cells over 2010. We repeat this process for simulated experimental periods 2011 and 2012, training the model on the prior three years of data. We display the parameter values in Table 1 and the ROC-like accuracy curve $f(x)$ aggregated over the three simulated experimental periods 2010, 2011, and 2012 in Figure 1.

The temporal and spatial length scale parameters in Table 1 have consistent values across the three model training periods. The spatial length scale parameter σ has an average estimated value of .0014 in the geographic coordinate system, which corresponds to approximately $165m$. The estimated time

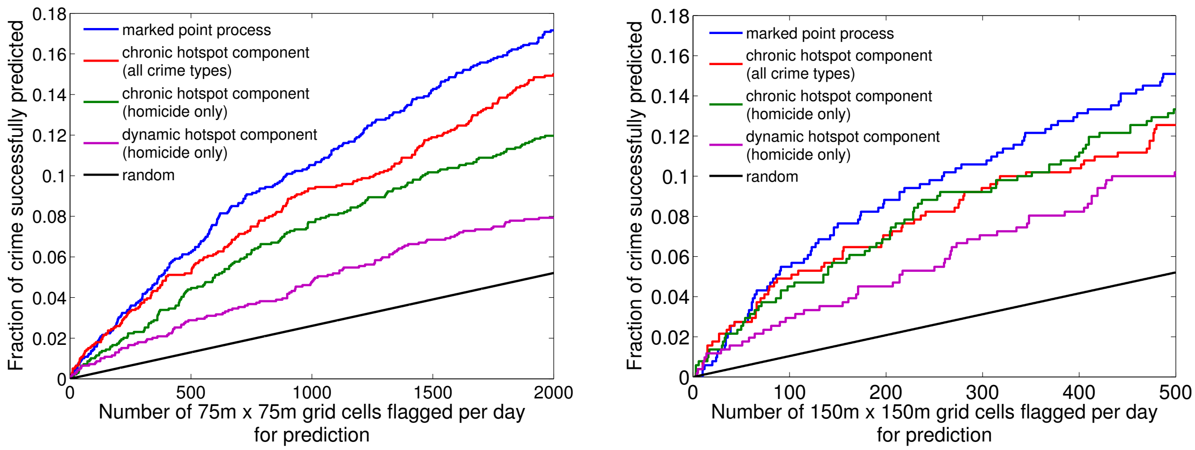

Figure 1: Fraction of homicide crime predicted over 2010-2012 versus number of grid cells flagged for intervention each day.

	ω	σ	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6
'07-'09	.024	.0015	.0046	.0001	.0022	.0068	.0220	.0000
'08-'10	.026	.0012	.0053	.0011	.0003	.0065	.0183	.0000
'09-'11	.023	.0016	.0093	.0000	.0007	.0025	.0311	.0000

	α_1	α_2	α_3	α_4	α_5	α_6
'07-'09	.1242	.0094	.0279	.0383	.0675	.0000
'08-'10	.0710	.0127	.0446	.0345	.0541	.0000
'09-'11	.0960	.0138	.0232	.0457	.0447	.0257

Table 1: Parameter estimates for mixture model (7)-(8) fit to Chicago gun crime data. Parameters are estimated over three different training periods: 2007-2009, 2008-2010, and 2009-2011.

scale parameter ω for the dynamic component of the point process corresponds to 41 days. These values are consistent with time and spatial bandwidths used in (Bowers et al., 2004; Mohler et al., 2011). Over the three training periods, the amount of crime attributed to the dynamic component of the model varies between 12-15%.

In Table 2, we display the percentage contributions of each crime type to the overall estimated intensity. Consistent with past research on leading indicators (Cohen et al., 2007), non-homicide crime types play an important role in the prediction of future homicide. Homicide components of the model only account for 8-13% of the overall intensity. The addition of leading indicator crime types serves to reduce the variance in the estimate of the intensity of homicide.

We compare the overall accuracy of the marked point process in Figure 1 to the accuracy of the various components of the model. For example, the chronic hotspot component given by Equation (8) is analogous to chronic hotspot policing methods (Weisburd et al., 2012) that fix hotspots for the entirety of the experimental period. The dynamic hotspot component given by the second term in Equation (6) is analogous to methods such as Prospective Hotspotting (Bowers et al., 2004) that capture near-repeat

	Homicide	Robbery	Assault	Weapons Viol.	Battery	Sex. Assault
'07-'09	13%	12%	16%	26%	33%	0%
'08-'10	8%	17%	24%	24%	27%	0%
'09-'11	11%	18%	13%	29%	29%	1%

Table 2: Contributions of different crime types to the estimated intensity of homicide over the three model training periods.

patterns. Because the marked point process is an off-grid method and the introduction of a grid adds a second layer of smoothing, we provide accuracy curves for both 150m×150m and 75m×75m discretizations of Chicago. Larger cell sizes obscure the accuracy of the point process method, whereas small cell sizes, such as 10m on a side, make hotspot policing impractical.

In Figure 1, the accuracy of the marked point process (all components) is the highest, followed by the chronic hotspot component (all crime types), chronic component (homicide only), dynamic component (homicide only). We believe that the order in terms of accuracy of the individual model components is due to variance reduction provided by adding more data (though for 150m grid cells the homicide only chronic component performs similarly to the chronic component with all crime types, as larger grid cells tend to obscure model performance). All methods significantly improve upon random predictions. While the dynamic component has the lowest accuracy, it's addition to the marked point process leads to a significant accuracy improvement over the chronic hotspot map that is fixed for the duration of the experimental periods. For 150m x 150m cells, the marked point process corresponds to a 10-20% accuracy increase over the chronic hotspot map for 1-500 grid cells flagged each day (all methods converge to the same accuracy as the number of flagged cells increases toward the number of cells in Chicago). To give some perspective on how many flagged cells are realistic, in six-month predictive policing experiments in Los Angeles twenty 150m×150m cells were flagged each day in each of three Los Angeles policing divisions. As Los Angeles has 21 divisions, this would correspond to 420 cells flagged each day city wide for directed patrol. While the geographic size of Los Angeles is larger than Chicago (1290 km^2 to 606 km^2), Chicago Police Department has more sworn officers (12,244 to 10,023). Thus a range of 250-500 150m×150m cells is a realistic number for a city the size of Chicago. For example, at 400 150m×150m cells flagged per day over the three year period, 180 homicides fall within the prediction cells of the marked point process, compared to 153 for the chronic hotspot map, a 17% improvement in accuracy. Given the high societal cost associated with homicide and serious gun crime (DeLisi et al., 2010), estimated to be in the billions of dollars a year in Chicago, even a few percent decrease in the homicide rate due to hotspot policing with a more accurate ranking method would be of significant societal benefit.

6. Conclusion

Hotspot and predictive policing strategies make possible the efficient distribution of limited policing resources and help police departments achieve crime rate reductions. Developments in mathematical and statistical modeling, high performance computing, and GPS enabled mobile devices make it possible for real-time crime forecasts to be at the disposal of officers in the field. Hotspot maps have their place in these technological solutions, yet in practice they often utilize too little of the data police agencies collect or fail to capture short-term changes in risk. In this article we showed how these methods can be improved by combining short-term and long-term kernel density estimates. We presented a marked point process model that systematically incorporates multiple crime types and several years of crime data into crime hotspot maps. The model can be estimated efficiently using an EM algorithm and thus can be easily deployed on a desktop computer or as a cloud based solution connected to a police agency RMS.

In future work, accuracy might be further improved by incorporating crime attractors into the density estimate of μ. A third kernel g_3 could be centered at each crime attractor location and estimated along with the rest of the model using the EM algorithm. Another way to improve model performance would be to more closely optimize the objective function. Maximum Likelihood penalizes model inaccuracy in low and medium risk areas of the city, whereas only the accuracy in the top ranked hotspots is of interest to police. Because the objective function is non-smooth due to the ranking step, learning to rank algorithms may be well suited for short-term crime prediction. Finally, retrospective accuracy assessments like the one in this paper should be followed up by experimental field trials to assess model performance within the context of predictive policing.

References

[1] Bowers, K. J., Johnson, S. D., & Pease, K. (2004). Prospective Hot-Spotting: The Future of Crime Mapping? *Brit. J. Criminol.*, **44**, 641–658.

[2] Braga, A. (2001). The Effects of Hot Spots Policing on Crime, *Annals of the American Academy of Political and Social Science*, **578**, 104–125.

[3] Chainey, S., Tompson, L., & Uhlig S. (2008). The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime, *Security Journal*, **21**, 4–28.

[4] Clarke, R. & Eck, J. (2005). Crime analysis for problem solvers in 60 small steps.

[5] Chainey, S. & Ratcliffe, J. (2005). *GIS and Crime Mapping*, West Sussex: Wiley.

[6] Cohen, J., W.L. Gorr, & A. Olligschlaeger (2007) Leading Indicators and Spatial Interactions: A Crime Forecasting Model for Proactive Police Deployment, *Geographical Analysis*, Vol. 39, Issue 1, pp. 105-127.

[7] DeLisi, M., Kosloski, A., Sween, M., Hachmeister, E., Moore, M., & Drury, A. (2010). Murder by numbers: Monetary costs imposed by a sample of homicide offenders, *The Journal of Forensic Psychiatry & Psychology*, 21(4), 501-513.

[8] Farrell, G. & Pease, K. (Ed.) (2001). *Repeat Victimization*, Criminal Justice Press, New York.

[9] Gorr, W. L. & Y. J. Lee (2012), A Longitudinal Study of Crime Hot Spots: Dynamics and Impact on Part 1 Violent Crime, Proceedings of the 32nd International Symposium on Forecasting.

[10] Gorr W. L. (2009) Forecast accuracy measures for exception reporting using receiver operating characteristic curves. International Journal of Forecasting 25: 48–61.

[11] Johnson, S. D. (2008). Repeat burglary victimisation: a tale of two theories, *Journal of Experimental Criminology* **4**, 215–240.

[12] Johnson, S. D., Bernasco, W., Bowers, K. J., Elffers, H., Ratcliffe, J., Rengert, G., & Townsley, M. (2007). Space-time patterns of risk: A cross national assessment of residential burglary victimization, *J. Quant. Crim.*, **23**, 201–219.

[13] Kennedy, L., Caplan, J. & Piza, E. (2010). Risk Clusters, Hotspots, and Spatial Intelligence: Risk Terrain Modeling as an Algorithm for Police Resource Allocation Strategies, *J. Quant. Crim.*, **27** (3), 339–362.

[14] Liu, H. & Brown, D. (2003). Criminal incident prediction using a point-pattern-based density model, International Journal of Forecasting, **19**, 603–622.

[15] Marsan D. & Lenglin, O. (2008). Extending Earthquakes Reach Through Cascading, *Science*, **319**, 1076–1079.

[16] Mohler, G., Short, M., Brantingham, P., Schoenberg, F. & Tita, G. (2011). Self-exciting point process modeling of crime, *JASA*, **106** (493), pp 100-108.

[17] Murphy, Kevin. (2012). Machine Learning: A Probabilistic Perspective. Cambridge, The MIT Press.

[18] Short, M.B. *et al.*.. (2009). Measuring and Modeling Repeat and Near-Repeat Burglary Effects, *J. Quant. Criminol.* **25** (3), 325.

[19] Veen, A. & Schoenberg, F. (2008). Estimation of space-time branching process models in seismology using an EM-type algorithm, *JASA*, **103** (482), pp 614–624.

[20] Wang, X. & Brown, D. (2012). The spatial-temporal modeling for criminal incidents, *Security Informatics*, **1** (1), DOI: 10.1186/2190-8532-1-2.

[21] Wang, X., Brown, D., & M. Gerber. Spatio-temporal modeling of criminal incidents using geographic, demographic, and twitter-derived information. *Intelligence and Security Informatics*, IEEE International Conference on. IEEE, 2012.

[22] Weisburd, D., Groff, E., & Yang, S.-M. (2012). The Criminology of Place: Street Segments and Our Understanding of the Crime Problem. Oxford, Oxford University Press.

[23] Zhuang, J., Ogata, Y., & Vere-Jones, D. (2002). Stochastic Declustering of Space-Time Earthquake Occurences, *JASA*, **97** (458), 369–380.